DEVELOPMENTS IN STATISTICS

VOLUME 1

EDITED BY

PARUCHURI R. KRISHNAIAH

Developments in Statistics

Edited by *PARUCHURI R. KRISHNAIAH*
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF PITTSBURGH
PITTSBURGH, PENNSYLVANIA

Volume 1

ACADEMIC PRESS New York San Francisco London 1978
A Subsidiary of Harcourt Brace Jovanovich, Publishers

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DX

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 77-11215

ISBN 0-12-426601-0

PRINTED IN THE UNITED STATES OF AMERICA

List of Contributors

Numbers in parentheses indicate the pages on which the authors' contributions begin.

A. V. BALAKRISHNAN (1), System Science Department, University of California at Los Angeles, Los Angeles, California

DAVID R. BRILLINGER (33), Department of Statistics, University of California at Berkeley, Berkeley, California and Department of Mathematics, University of Auckland, Auckland, New Zealand

P. R. KRISHNAIAH (135), Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

M. M. RAO (171), Department of Mathematics, University of California at Riverside, Riverside, California

PRANAB KUMAR SEN (227), Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, North Carolina 27514

J. N. SRIVASTAVA (267), Department of Statistics, Colorado State University, Fort Collins, Colorado 80523

Preface

The series "Developments in Statistics" has been created to provide a central medium for the publication of long and important papers in various branches of statistics. The papers may be (i) expository papers, (ii) research papers, or (iii) papers that are partially expository in nature. The volumes in the series will appear at irregular intervals. The papers in these volumes are, in general, too long to be published in journals but too short to be published as separate monographs. The series will cover both theory and applications of statistics. The first volume consists of invited papers written by outstanding workers in the field. These papers give authoritative reviews of the present state of the art on various topics, including new material in many cases, in the general areas of stochastic control theory, point processes, multivariate distribution theory, time series, nonparametric methods, and factorial designs.

I wish to thank the Department of Mathematics and Statistics at the University of Pittsburgh and the Department of Statistics at Carnegie-Mellon University for providing the facilities to edit this volume. I also wish to thank Academic Press for its excellent cooperation.

Contents of Volume 2

Contents

Covariance Analysis of Nonstationary Time Series

M. M. RAO

Nonparametric Repeated Significance Tests

PRANAB KUMAR SEN

A Review of Some Recent Work on Discrete Optimal Factorial Designs for Statisticians and Experimenters

J. N. SRIVASTAVA

Parameter Estimation in Stochastic Differential Systems: Theory and Application

A. V. BALAKRISHNAN†

SYSTEM SCIENCE DEPARTMENT
UNIVERSITY OF CALIFORNIA AT LOS ANGELES
LOS ANGELES, CALIFORNIA

1. INTRODUCTION

The estimation problem in essence is the following. We have an observed process $y(t)$ ($n \times 1$ matrix function) which has the form

$$y(t) = S(\theta, t) + N(t), \qquad 0 < t < T \tag{1.1}$$

where θ denotes a vector of unknown parameters which we want to estimate, $S(\theta, t)$ being a stochastic process (*signal*) which is completely specified once θ is specified (e.g., by means of a stochastic differential system) and $N(t)$ is a stochastic process which models the errors (that remain even after all

† Research was supported in part by Grant No. 73-2492, Applied Mathematics Division, AFOSR, USAF.

systematic errors, such as bias and calibration errors, have been accounted for). There is much evidence to suggest that the noise process may be well modeled as Gaussian, and independent of the signal process. This is a basic assumption throughout this chapter.

Under the title of "system identification" there is a large engineering literature dealing with such problems. This is well documented in the proceedings of three symposia [1] devoted exclusively thereto. In the bulk of this literature, the process $S(\theta, t)$ is taken to be deterministic, in which case the estimation is largely treated as a *least squares* problem of minimizing

$$\int_0^T \|y(t) - S(\theta, t)\|^2 \, dt$$

over a predetermined admissible set of parameters θ. Where the stochastic signal case is considered, it is reduced to the time-discrete version of (1.1):

$$y_n = S_n(\theta) + N_n \tag{1.2}$$

for the reason that the continuous time is mathematically too difficult to handle, and anyhow, in digital computer processing (as is the rule), it is so discretized in the analog-to-digital (A–D) conversion process. This is indeed true; but the authors invariably proceed to make the assumption that the noise samples $\{N_n\}$ are mutually independent. But this requires that the sampling rate (in the periodic sampling of the data) be not more than twice the postulated *bandwidth* of the noise, itself actually unknown. Indeed in most practical cases the sampling rate is far higher than twice the bandwidth. To meet this objection, one may then allow the $\{N_n\}$ to be correlated. But then the correlation function must be known, and anyone with experience in handling real data can easily appreciate that it is unrealistic to require that much knowledge of the noise process, even if the complication in the theory can be borne.

We maintain, in any event, that it is much better to work with the time-continuous model (1.1), allowing as high a sampling rate in the processing as the A–D converter is designed for. But in the time-continuous model we are faced with another problem. The basic tool in estimation is the likelihood functional (for fixed parameters) which is based on the Radon–Nikodym derivative of the probability measure induced by the process by $y(\cdot)$ to that induced by the noise process $N(t)$. But this derivative is too difficult to calculate even when the precise spectrum of $N(\cdot)$ is known, which it is not. What we can assert for sure is that the bandwidth of noise $N(t)$ is much larger than that of the process $S(\theta, t)$, which is essential in order that the measuring instrument does not *distort* the signal. At this point it was customary in the earlier engineering literature to introduce *white noise* in a formal way as a stationary stochastic process with constant spectral density to represent the *large bandwidth* nature of $N(t)$. With the advances in the theory of diffusion processes using the Ito integral, it became fashionable to use a Wiener process model as being *more rigorous* [2]. Thus we replace (1.1) by

$$Y(t) = \int_0^t S(\theta, \sigma)\, d\sigma + W(t) \tag{1.3}$$

where $W(t)$ is a Wiener process. We can then exploit the well-developed machinery of martingales and Ito integrals. In fact the likelihood function can then be expressed as (see Liptser and Shiryayev [2])

$$\exp\left\{-\tfrac{1}{2}\int_0^T \|\hat{S}(\theta, t)\|^2\, dt - 2\int_0^T [\hat{S}(\theta, t), dY(t)]\right\} \tag{1.4}$$

where $\hat{S}(\theta, t)$ is the best mean square estimate of $S(\theta, t)$ given the σ-algebra generated by $Y(s)$, $s \le t$. This formula can be justly considered as one of the triumphs of the Ito theory, the key to the success being the appearance of the Ito integral in the second form of (1.4). This integral is defined on the basis that $Y(t)$ is of unbounded variation with probability 1. Of course no physical instrument can produce such a waveform. To calculate it, given the actual observation (1.1), we can "retrace" our steps back from (1.3) and use $y(t)\, dt$ in place of $dY(t)$. But this is totally incorrect, unless $S(\theta, t)$ is deterministic, and any minimization procedure based on it leads to erroneous results. This point is not appreciated by authors using (1.3) as more rigorous, perhaps because they have not had occasion actually to calculate anything based on real data. In any data generated by digital computer simulation, which must perforce employ the discrete version (1.2), this point can be completely masked and hence never appreciated.

Faced with this difficulty we have to examine more precisely the model again, to see a physically more meaningful way of exploiting the fact that the noise bandwidth is large compared to the signal bandwidth. What is needed is the *asymptotic form* of the likelihood functional as the bandwidth expands to infinity in an arbitrary manner.

Such a theory has been developed by the author using a precise notion of white noise. This is explained in Section 2. Based on this theory we derive a likelihood functional in Section 3. It turns out that formula (1.4) is replaced by

$$\exp\left\{-\tfrac{1}{2}\int_0^T \|\hat{S}(\theta, t)\|^2\, dt - 2\int_0^T \hat{S}(\theta, t) y(t)\, dt \right.$$
$$\left. + \int_0^T \left(\widehat{\|S(\theta, t)\|^2} - \|\hat{S}(\theta, t)\|^2\right) dt\right\} \tag{1.5}$$

where the caret denotes conditional expectation given the data up to time t. Note that a third term appears, which can also be expressed as

$$\int_0^T \|S(\theta, t) - \hat{S}(\theta, t)\|^2 \, dt$$

and in the case in which $S(\theta, t)$ is Gaussian, this reduces to

$$\int_0^T E[\|S(\theta, t) - \hat{S}(\theta, t)\|^2] \, dt$$

being thus the integral of the mean square error in estimation of the signal $S(\theta, t)$ from the observation up to time t. When the signal process can be described in terms of stochastic differential equations, whether finite or infinite dimensional, advantage can be taken of the fact that the mean square error can be evaluated by solving a Riccati equation. Section 4 is devoted to this specialization. Section 5 deals with the application to the problem of stability and control derivatives from flight test data taking turbulence into account. The algorithms used and results obtained on actual flight data are included.

2. WHITE NOISE: BASIC NOTIONS

Let H denote a real separable Hilbert space and let

$$W = L_2[0, T; H], \qquad 0 < T < \infty$$

denote the real Hilbert space of H-valued weakly measurable functions $u(\cdot)$ such that

$$\int_0^T [u(t), u(t)] \, dt < \infty$$

with the inner product defined by

$$[u, v] = \int_0^T [u(t), v(t)] \, dt$$

Let μ_G denote the Gaussian measure on W (on the cylinder sets with finite-dimensional Borel basis) with characteristic function

$$C_G(h) = \exp(-\tfrac{1}{2}[h, h]), \qquad h \in W$$

Elements of W under this (finitely additive) measure will be *white noise sample functions*, denoted by ω. This terminology appears to have the sanction of usage; see Skorokhod [3] for example. It is essential for us that W is an L_2-space over a finite interval.

Any function $f(\cdot)$ defined on W into another Hilbert space H_r such that the inverse images of Borel sets in H_r are cylinder sets with base in a finite-dimensional subspace will be called a *tame* function (see Gross [4]). As is readily seen, the class of tame functions is a linear class. Since the inverse image of the whole space H_r must be cylindrical, it is clear that any tame function has the form $f(P\omega)$ where P is a finite-dimensional projection.

To introduce the notion of a *random variable* let us first confine ourselves to the case in which H_r is finite dimensional: $H_r = R^n$ say. We introduce a metric into the linear space of tame functions by

$$|||f-g||| = \int_W \frac{\|f-g\|}{1+\|f-g\|}\, d\mu_G$$

and then complete the space, the completion yielding a Frechet space. Every element of the completed space is called a random variable and if ζ denotes such an element and $f_n(\omega)$ a corresponding Cauchy sequence in probability, then we define the corresponding *distribution function* or probability measure on R^n to be that induced by the characteristic function

$$C_\zeta(h) - \lim_n E(\exp\{i[f_n(\omega), h]\}) \tag{2.0}$$

The latter limit exists (uniformly on bounded sets of $R^n = H_r$).

In the case in which H_r is no longer finite dimensional, we shall still identify Cauchy sequences in probability of tame functions as *weak random variables.* The limit in (2.0) still holds uniformly on bounded sets in H_r, but the limit may in general only define a *weak distribution* on H_r. We recall in this connection the Sazonov theorem [5] that the limit is the characteristic function of a probability measure if and only if it is continuous in the trace-norm topology (*S-topology,* see later). This is automatically the case if the sequence is Cauchy in the mean square sense, and we shall then drop the qualification weak.

Let $f(\omega)$ be any Borel measurable function mapping W into H_r. Then $f(P\omega)$ is tame for every finite-dimensional projection operator P. Let $\{P_n\}$ denote a sequence of finite-dimensional projections converging strongly to the identity; the sequence may be assumed to be monotone. If the sequence $f(P_n\omega)$ is Cauchy in probability, then we may associate a (weak, in general) random variable with $f(\cdot)$. Let us denote it by $f^{\sim}$ (a notation used by Gross). This limit of course can depend on the particular projection sequence chosen. Of primary interest to us are those function $f(\cdot)$ for which $\{f(P_n\omega)\}$ is Cauchy in probability for *every* such sequence of finite-dimensional projections, and moreover such that all such Cauchy sequences are equivalent so that the limit random variable $f^{\sim}$ is unique. In that case we say that $f(\omega)$ is a (weak) random variable. We shall use the term random variable if the corresponding measure is countably additive; with mean square convergence, this will be automatic.

The simplest function one can consider is perhaps the linear function

$$f(\omega) = L\omega$$

where L is a linear bounded transformation mapping W into H_r, where we now allow H_r to be infinite dimensional. Then it is easy to see that if L is Hilbert–Schmidt, then $\{LP_m\omega\}$ is Cauchy in the mean square sense, and $L\omega$ is a random variable. Conversely, L must be H–S if $L\omega$ is to be a random variable.

What is the class of functions which are random variables? To answer this question, at least in part, let us introduce the S-topology on W: This is the (locally convex) topology induced by seminorms of the form

$$\rho(\omega) = [S\omega, \omega]^{1/2} \tag{2.1}$$

where here (and hereinafter) S denotes a self-adjoint, nonnegative definite trace-class operator on W into W. For the case in which $H_r = R^1$, Gross [4] has given a sufficient condition: $f(\cdot)$ is a random variable if it is *uniformly* continuous in the S-topology. Uniform continuity means that given $\varepsilon > 0$, we can find $\rho(\cdot)$ such that

$$\|f(x) - f(y)\| < \varepsilon \qquad \text{for all} \quad x, y \quad \text{such that} \quad \rho(x-y) < 1$$

Unfortunately Gross does not seem to discuss nontrivial examples of functions satisfying this condition. Here we shall give a sufficient condition for a class of random variables with finite second moment.

Theorem 2.1. Let $p_n(\omega)$ denote a homogeneous polynomial of degree n mapping W into H_r. Suppose it is continuous at the origin in the S-topology. Let P denote any finite-dimensional projection. Then

$$\sup_P E(\|p_n(P\omega)\|^2) < \infty \tag{2.2}$$

where the supremum is taken over the class of all finite-dimensional projections. Conversely, if (2.2) holds, then $p_n(\cdot)$ is continuous at the origin in the S-topology.

Proof. We begin with a simple but useful lemma.

Lemma 2.1. Suppose $p_n(\cdot)$ is continuous in the S-topology at the origin. Then there exists a seminorm in the S-topology:

$$\rho(\omega) = [S\omega, \omega]^{1/2} \tag{2.3}$$

such that

$$\|p_n(\omega)\| \leq M\rho(\omega)^n \tag{2.4}$$

where M is a constant. Conversely, if (2.4) holds, then $p_n(\omega)$ is continuous in the S-topology at the origin.

Proof. Continuity in the S-topology at zero implies the following: Given $\varepsilon > 0$ we can find a seminorm of the form (2.3) such that

$$\|p_n(\omega)\| < \varepsilon \qquad \text{for all} \quad \omega \quad \text{such that} \quad \rho(\omega) \leq \delta \tag{2.5}$$

Hence for any ω for which $\rho(\omega) \neq 0$, we have that

$$\left\| p_n \left(\frac{\delta\omega}{\rho(\omega)} \right) \right\| < \varepsilon$$

or by the homogeneity of $p_n(\cdot)$,

$$\|p_n(\omega)\| < \left(\frac{\varepsilon}{\delta^n} \right) \rho(\omega)^n, \qquad \rho(\omega) \neq 0$$

If $\rho(\omega) = 0$, then for any positive number k,

$$\rho(k\omega) = 0$$

and hence from (2.5)

$$\|p_n(\omega)\| < \varepsilon \,|\, k^n \qquad \text{for all} \quad k > 0$$

and hence

$$p_n(\omega) = 0$$

Therefore (2.4) holds. The converse is obvious.

Proof of Theorem. Corresponding to a finite-dimensional projection P, we can find an orthonormal basis $\{\phi_i\}$ such that P is the projection operator corresponding to the space spanned by ϕ_i, $i = 1, 2, \ldots, m$. Let

$$p_n(\omega) = k_n(\omega, \ldots, \omega)$$

$k_n(\cdots)$ being the symmetric n-linear form, corresponding to $p_n(\cdot)$. Then

$$p_n(P\omega) = \sum_{i_1=1}^{m} \cdots \sum_{i_n=1}^{m} a_{i_1, \ldots, i_n} \zeta_{i_1}, \ldots, \zeta_{i_n} \tag{2.6}$$

where

$$a_{i_1, \ldots, i_n} = k_n(\phi_{i_1}, \ldots, \phi_{i_n}), \qquad \zeta_i = [\phi_i, \omega]$$

$\{\zeta_i\}$ is a sequence of independent zero-mean unit variance Gaussians and

(2.6) defines a tame function. Moreover we can readily calculate [by expressing (2.6) in terms of Hermite polynomials for instance] that

$$E(\|p_n(P\omega)\|^2) = \sum_{\nu=0}^{[n/2]} \left(\frac{n!}{(n-2\nu)!\,2^\nu \nu!}\right)^2 \times \sum_{i_{2\nu+1}=1}^{m} \sum_{i_n=1}^{m} \left\| \sum_{i_1=1}^{m} \cdots \sum_{i_\nu=1}^{m} a_{i_1, i_1, \ldots, i_\nu, i_\nu, i_{2\nu+1}, \ldots, i_n} \right\|^2 \tag{2.7}$$

But from Lemma 2.1, we have that

$$\|p_n(P\omega)\|^2 \le [S_m\omega, \omega]^n \tag{2.8}$$

where $S_m = PSP$, and is of course trace-class and finite dimensional. Hence

$$E[\|p_n(P\omega)\|^2] \le E([S_m\omega, \omega]^n) \tag{2.9}$$

Let ψ_k, $k = 1, \ldots, \nu$, be the orthonormalized eigenvectors of S_m with corresponding nonzero eigenvalues λ_k. Then

$$[S_m\omega, \omega] = \sum_1^\nu \lambda_i[\psi_i, \omega]^2$$

and we have

$$E([S_m\omega, \omega]^n) = f(\text{tr } S_m, \text{tr } S_m{}^2, \ldots, \text{tr } S_m{}^n)$$

where $f(\cdot)$ is a fixed continuous function. Of course tr $S_m{}^j$ is monotone in m for each j and converges to tr S^j. Hence it follows that

$$E[\|p_n(P\omega)\|^2] < \infty$$

for *all* finite-dimensional projections.

To prove the converse, suppose (2.2) holds. Then (2.7) holds for every m, and taking $\nu = 0$ therein, we obtain that

$$\sum_{i_1}^{\infty} \cdots \sum_{i_r}^{\infty} \|k_n(\phi_{i_1}, \ldots, \phi_{i_n})\|^2 < \infty \tag{2.10}$$

for every orthonormal sequence $\{\phi_i\}$. Hence $p_n(\cdot)$ is Hilbert–Schmidt. Of course

$$\|p_n(\omega)\|^2 \le M\|\omega\|^{2n} \tag{2.11}$$

Define S now by

$$[S\omega, \omega] = (\|p_n(\omega)\|^2)^{1/n}$$

Then S is Hilbert–Schmidt by (2.10). For any finite-dimensional projection P,

$$E[SP\omega, P\omega] = E[PSP\omega, \omega] = E((\|p_n(P\omega)\|^2)^{1/n})$$